A survey of multimodal deep generative models

Year: 2022 | Citations: 104 | Authors: Masahiro Suzuki, Y. Matsuo

Abstract

Multimodal learning is a framework for building models that make predictions based on different types of modalities. Important challenges in multimodal learning are the inference of shared representations from arbitrary modalities and cross-modal generation via these representations; however, achieving this requires taking the heterogeneous nature of multimodal data into account. In recent years, deep generative models, i.e. generative models in which distributions are parameterized by deep neural networks, have attracted much attention, especially variational autoencoders, which are suitable for accomplishing the above challenges because they can consider heterogeneity and infer good representations of data. Therefore, various multimodal generative models based on variational autoencoders, called multimodal deep generative models, have been proposed in recent years. In this paper, we provide a categorized survey of studies on multimodal deep generative models. GRAPHICAL ABSTRACT